

Original Articles

Evaluation of multispectral data for recent manure application: A case study in northern Spain

Oscar D. Pedrayes^a, Rubén Usamentiaga^{a,*}, Yanni Trichakis^b, Faycal Bouraoui^b

^a Department of Computer Science and Engineering, University of Oviedo, Campus de Viesques, Gijón 33204, Asturias, Spain

^b European Commission – Joint Research Centre, Via Enrico Fermi 2749, Ispra 21027, Varese, Italy



ARTICLE INFO

Keywords:

Remote sensing
Multispectral indices
Pasture
Machine learning
Leaching
Manure

ABSTRACT

The use of manure in agricultural fields during the wet season can lead to environmental pollution by releasing nitrates into nearby water sources. To address this issue, authorities may impose closed periods during which manure application is prohibited. However, ensuring compliance with these regulations can be challenging, as it is difficult to monitor all fields in a country. To tackle this problem, a solution has been proposed that involves employing machine learning techniques in conjunction with satellite imagery to automatically identify freshly manured fields. This paper investigates the relationship and effectiveness of the Sentinel-2 satellite bands and 51 frequently utilized multispectral indices in the context of precision agriculture, by exploring different feature selection methods. The proposed method achieves nearly 90% F₁-Score and detects all test plots of the northern Spanish region, showing its potential for large-scale use in precision agriculture and environmental monitoring. This method incorporates temporal data, resulting in an 8% improvement in the detection F₁-Score. Despite their lower spatial resolution, infrared bands have proven to be more effective than visible bands, enhancing the F₁-Score by 4%. Furthermore, the use of over 80 features contributes to a 12% increase in the F₁-Score compared to using fewer than 10 features. For further research and future studies, a dataset of recently manured plots, verified on-site, has been developed and made publicly available.

1. Introduction

Multispectral remote sensing (Colwell, 1966) is a powerful technology that captures image data within specific wavelength ranges throughout the electromagnetic spectrum. This technology offers the unique capability to distinguish different features based on their spectral responses in various bands of the spectrum. The data collected from these bands can be utilized individually or combined to form multispectral indices. These indices are constructed to enhance the contribution of certain features while minimizing the influence of others, providing more precise information about specific phenomena such as vegetation conditions, soil type, and temperature (Curran, 1980). In this way it is possible to highlight different phenomena in the images that would normally be more difficult to observe, extending the capabilities of the bands.

Multispectral indices exist for a wide variety of subjects. In particular, in this publication, multispectral indices of the precision agriculture literature are studied. These indices are called vegetation indices

and are used for a wide variety of tasks ranging from the detection of vegetation and its health, to the detection of specific substances such as nitrates or nitrogen. Looking at these applications, the possibility of detecting the spreading of manure on crops arises. This can be of great importance for environmental conservation as nitrates from manure, as well as other elements from supplements given to livestock (Brugger and Windisch, 2015), can reach into nearby water bodies contaminating them (Tzilivakis et al., 2021; Liu et al., 2018; Kleinman et al., 2020). This phenomenon is known as leaching and is accentuated in rainy seasons. For example, there are recent reports about the death of large quantities of fish due to the lack of oxygen caused by leaching contamination (Kirchman, 2022; Fahmy, 2022; Gillespie, 2022). For this reason, legislations are created to prohibit (Council of the European Union, 1991; Colwell, 2000) or regulate (Ironmonger, 2022; Bouma, 2016) the use of manure as fertilizer during these periods. This is formally known as closed periods and can prevent environmental pollution (Tzilivakis et al., 2021). Closed periods may vary depending on the vulnerability of the terrain and other risks, such as the climate of that

* Corresponding author.

E-mail addresses: oscardiazpedrayes@gmail.com (O.D. Pedrayes), rusamentiaga@uniovi.es (R. Usamentiaga), Ioannis.TRICHAKIS@ec.europa.eu (Y. Trichakis), Faycal.BOURAOUI@ec.europa.eu (F. Bouraoui).

<https://doi.org/10.1016/j.ecolind.2024.111550>

Received 20 July 2023; Received in revised form 25 November 2023; Accepted 1 January 2024

Available online 9 January 2024

1470-160X/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

particular region. A solution based on multispectral indices from satellite data can enable monitoring and enforcement of the laws, as on-site investigations are unfeasible given the scale.

Precision agriculture using satellite imagery is a well-researched field. There is plenty of literature about crop classification (Orynbai-kyzy et al., 2019; Pedrayes et al., 2021), or the health of crops (Mutanga et al., 2017; Shanmugapriya et al., 2019). However, no existing literature explores the detection of recent manure application in different types of fields using a wide range of satellite multispectral indices. Most of the literature related to fertilizer detection studies only a few multispectral indices at a time, and uses aircraft imagery with much higher resolution (Yang et al., 2002; Qun'ou et al., 2021; Pereira et al., 2022). The problem is that aircraft imagery is expensive and cannot be used as a monitoring method. In the case of unmanned aerial vehicles (UAVs) such as drones, the cost of constantly monitoring large areas is not feasible. Both airplane and UAV imagery can provide high-resolution data and are useful for specific applications. However, they have limitations in terms of coverage area, flight time, and susceptibility to weather conditions such as wind, which may make satellite data a more practical option for certain applications (Alvarez-Vanhard et al., 2021; Analytics, 2022). Satellite data can offer global coverage, long-term data time-series, and environmental monitoring at different resolutions (Earth, 2019). In addition, some satellites offer their imagery for free. In the works that use satellite imagery to study vegetation, Sentinel-2 is the most common option in recent years (Ma et al., 2010; Romanko, 2017; Zhu et al., 2021; Dodin et al., 2021). This is because its data is publicly available for free and has 13 multispectral bands with a revisit time of only 5 days. Its spatial resolution is not particularly high with 10 to 30 meters; however, when compared to other satellites, it is among the best available options.

Most literature on multispectral indices for detection of fertilizers using Sentinel-2 satellite imagery are about visualization in the images (Wang, 2009), or study the correlation of the bands and some multispectral indices (Dodin et al., 2021). There is not extensive research on advanced machine learning classification for fertilizer detection and even less specifically for recent manure application. The literature agrees that the short-wave infrared bands and multispectral indices are the most representative for nitrates and nitrogen (Dodin et al., 2021; Romanko, 2017; Yang et al., 2002; Fu et al., 2021). There is some research for organic matter (Dodin et al., 2021; Wang, 2009) but this may include many different sources.

Another relevant study is (Shea et al., 2022). Its focus is on detecting soil moisture on clear days and distinguishing between saturated and dry soil conditions in agricultural fields. This is achieved through the use of median images created from multiple images taken over a 60-day period to avoid cloud cover. For this reason, this method is not intended for monitoring manure application. Instead, its goal is to identify spray fields based on saturated soil conditions. To do so, only periods without rainfall are used to ensure that no moisture is present, which further limits its usefulness as a monitoring method. Various techniques are employed to remove bodies of water such as lakes and rivers from the analysis. The detection of water is then used to infer the presence of liquid manure in sprayfields. However, the objective is not to detect illegal use of manure; in fact, the researchers attempt to exclude illegal uses and assume that all detected uses are legal (e.g., by discarding fields with slopes greater than 10 degrees). The dataset used is not publicly available and consists of only 1000 pixels with imbalanced crop types. It includes data from Sentinel-1, normalized difference vegetation index (NDVI) and modified normalized difference water index (MNDWI) from Sentinel-2 bands, thermal infrared (TIR) bands from Landsat 8, terrain elevation, and distance to lakes (as sprayfields are typically located near water sources to reduce transportation costs). For this reason, the data collected are not from the same days and do not consider several relevant multispectral indices. The reported recall and precision values are above 90%. However, it is unclear whether the separation of training and validation data is done at the plot or pixel level. Furthermore, the

dataset was not constructed through on-site verification but rather through visual inspection and assumptions (e.g., removing pixels with more than 90% humidity).

The most common classification methods using multispectral imagery are: decision trees (Yang et al., 2002) such as random forest (Zhu et al., 2021; Shea et al., 2022), or customized artificial neural networks (Jaihuni et al., 2021; Fu et al., 2021).

This topic may not be of interest to small companies or individuals, since in small regions it would not be necessary to use satellite imagery. Only a large entity such as a government of a region would be interested. On top of that, creating a sufficiently large dataset is time-consuming and costly because in situ investigations are necessary. This causes all the datasets in the literature to be private. For these reasons, and because it is a very specific topic, no public datasets containing satellite images of recently fertilized plots were found. In addition, most studies focus on a particular type of crop, such as wheat or maize. This may be because it is difficult to get a diverse dataset within the same region.

Although the low spatial resolution issue cannot be addressed directly, the analysis can be improved by exploring image time series. This is especially relevant in scenarios where the crops plots are small (Ma et al., 2010; Zhu et al., 2021). (Dodin et al., 2021) proposes to use multiple images of the same region at different times to create a time series and potentially improve detection. However, this approach can increase the difficulty when creating a dataset because some days are covered by clouds.

The objective of this paper is to find and evaluate the best set of multispectral indices for manure detection. As a result, it contributes to the reduction of the time and cost required to detect newly fertilized fields, which makes it possible to comply with closed-period laws. This study evaluates all Sentinel-2 bands as well as 51 of the most relevant multispectral indices in the literature for precision agriculture. Experiments using advanced machine learning methods using different sets of features are carried out to evaluate the multispectral indices. To select the feature sets, several state of the art feature selection methods are used. To increase available data about soil changes, multiple images per plot in different dates are used which increases the number of features. A new dataset consisting of 30 freshly manured fields is developed to train and test the models. To find fitting crop fields and to confirm that manure was recently applied, on-site investigations were conducted. This dataset is publicly available for future research.

This work represents a novel investigation into an aspect that has not previously been addressed in the literature, the evaluation of most relevant multispectral data for monitoring recent manure application using several feature selection and machine learning methods. The results of this study facilitates the detection of manure in fields, which can be further utilised in future research helping to ensure the enforcement of laws to reduce nitrate leaching in the environment.

2. Proposed approach

2.1. Methodology

To develop the proposed solution, the required training and test data must first be collected, its features must be extracted and studied, and finally the detection models must be trained. This section describes both the model-training and the detection pipeline.

2.1.1. Training pipeline

Fig. 1 depicts the sequence of steps required for training a model. The imagery to train the model must first be acquired from the Sentinel-2 satellite. On-site inspections are undertaken to pinpoint areas of interest to acquire the images. A manually pixel labeling procedure is used to select pixels for the ground truth mask necessary for model training from these areas of interest. To achieve as much variety as possible, parts of the image where the plots are not fertilized are chosen as counterexamples. These images have been edited to eliminate any areas that are

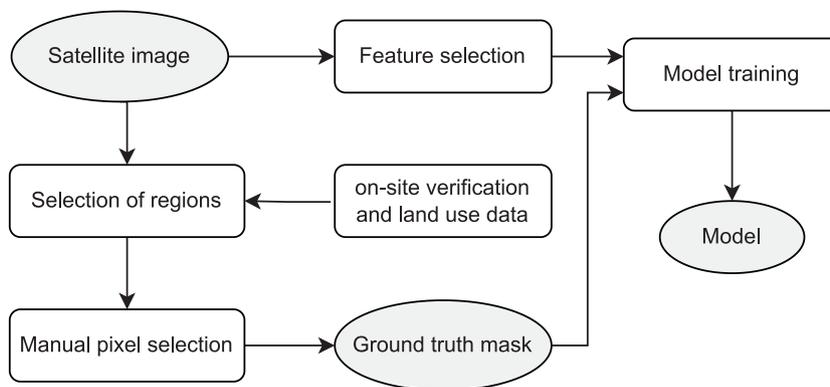


Fig. 1. Pipeline for model training.

not relevant to the study, such as areas that cannot be fertilized like highways, buildings, or forests. Land use data is gathered from an up to date national database to determine the areas of no interest. Due to their georeference, these pixels are automatically erased. The ground truth masks are prepared for use in the training procedure once this step is finished.

Each pixel’s features are calculated and defined in order to prepare the data for training. The bands can be used to calculate multispectral indices. In this study, various feature-based techniques are assessed. The initial image taken prior to the application of manure is also taken into account to maximize the amount of information. This inclusion effectively doubles the number of features available for analysis and provides valuable temporal information. These features are used to train the models. To prevent having pixels from the training and test sets adjacent to one another, plots rather than pixels are used to split the dataset. By doing this, the likelihood of overfitting is reduced. 70% of the plots that are available are used for training. The remaining 30% plots are used to evaluate the models’ performance. Common machine learning techniques such as Naive Bayes (Berrar, 2018), Decision Trees (Quinlan, 1986), Discriminant Analysis (Klecka, 1980), Logistic Regression (Hosmer et al., 2013), Support Vector Machines (Hearst et al., 1998), Nearest Neighbor (Cover and Peter, 1967), Kernels Approximation (Aizerman, 1964), Ensemble (Dietterich, 2000), and Neural Networks (Zupan, 1994) are evaluated into select the most effective method for the model training.

2.1.2. Detection pipeline

The pipeline necessary to implement the manure detection model is shown in Fig. 2. Obtaining the original image is a prerequisite, just as during training. In any case, it is always possible to use, in addition to the current image, an image from immediately before. This is because the system is intended for constant monitoring without the need to know the date of the manure. Regions that are not of interest, that is, regions having land uses unrelated to crops and fields, have been removed from

this image. This data was taken from a current national land use database. Features are computed and chosen using the remaining pixels. The generated feature image is loaded into the model to produce a binary mask indicating whether each associated pixel is part of a manure-applied plot. Morphological approaches are used for noise removal to visualize expected manure masks. These masks are frequently noisy because the detection is done pixel by pixel. This process is simply used for the final display and has no bearing on the metrics’ outcomes. Two methods are employed to remove mask noise (Serra, 1982): erosion, which reduces the size of the regions by erasing all floating pixels using a 2 × 2 square structure, and dilation, which restores the regions’ original size and fills in any gaps. By removing stray pixels and filling in the areas, the final masks are enhanced.

2.2. Image data acquisition

There is no publicly available data in the literature that can be applied to this investigation. As a result, a dataset was created manually. Sentinel-2 is the satellite that was selected for the imagery due to its spectral range, geographic resolution, and, most importantly, its 5-day revisit time. Specifically, the Level-2A product was used, which provides imagery with atmospheric correction.

To capitalize on the temporal information provided by Sentinel-2, it was chosen to use the image immediately before the application of manure, as well as the image immediately after. The difference between the two images might therefore provide substantial information about the soil and improve the results. This causes the number of features per pixel to be doubled.

2.3. Ground truth generation

Locating the target plot, creating its mask, removing irrelevant pixels such as roads or buildings, and then gathering counterexamples to

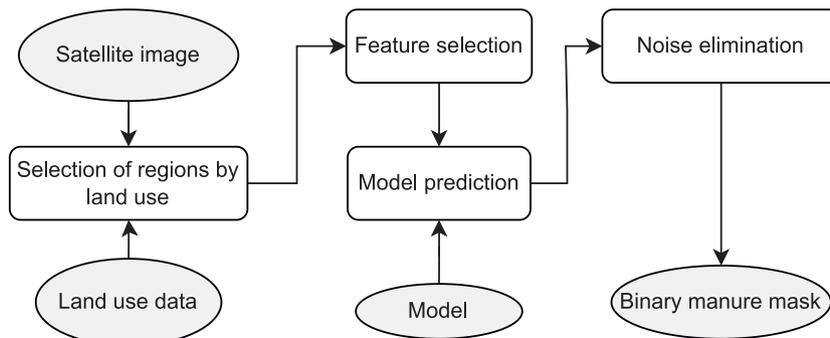


Fig. 2. Pipeline for model detection.

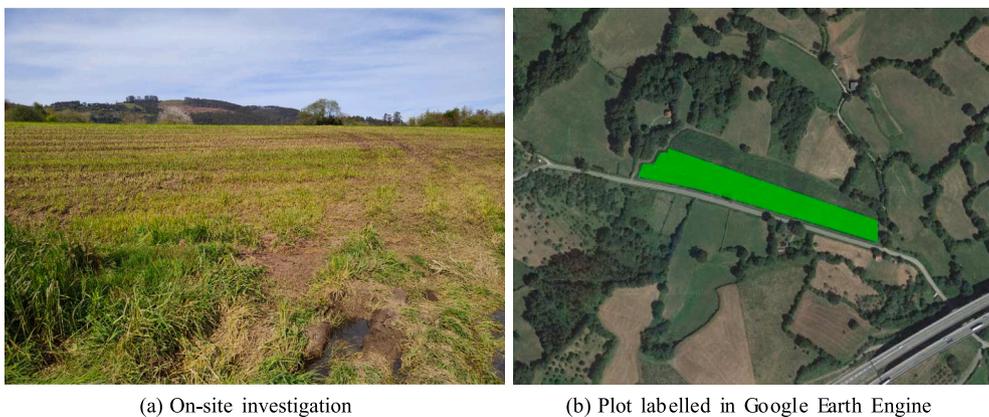


Fig. 3. P-VG1 labelling example.

properly train the models is required to generate a proper ground truth mask.

2.3.1. Plot localization

Once the plot has been identified, an on-site investigation is carried out to validate that it has recently been manured and to determine the real location of the manure in the field (see Fig. 3a). The plot is then labeled with the observed dimensions using Google Earth Engine, as seen in Fig. 3b. The Sentinel-2 georeferenced image is then utilized to construct the raster ground truth mask using the annotation, as illustrated in Fig. 4.

2.3.2. Regions of no interest

It is useful to delete pixels of no importance, such as highways, buildings, or bodies of water, to lessen the problem’s complexity. It is not important to take these places into account because they will not receive fertilization. A national land use database is used to complete this task reliably and automatically. Sentinel-2 provides a layer referred to as the “Scene classification layer,” but it is insufficient for this investigation because it cannot discriminate between classes like “Forest” and “Pastures.”.

2.3.3. Counterexamples

In order to train a classification model, it is crucial to have a counterexample class in which to categorize any pixels that do not fall under

the target class in addition to the data pertinent to the target class to be categorized. This implies that a classification requires the existence of at least two classes. The definition of this new class is “Others” class. “Manure application” class is the name given to the target class.

The counterexample pixels are manually selected from the complete images. Image regions in which the areas are verified to be non-fertilized are chosen. This selection includes visually similar soil such as plowed lands. To give the counterexamples additional context, certain areas with trees, roads, and buildings are also included. It is helpful to provide a few instances of these regions even though land use masks have already been removed in order to prevent border regions or regions that have been erroneously labelled from being mistakenly placed in the “Manure application” class.

2.4. Feature extraction

The features that were utilized to create the dataset are described in this Section 51 of the most popular multispectral indices mentioned in the literature for precision agriculture, together with the 13 bands of the Sentinel-2 satellite, are used as features in this study. Each pixel includes 64 features altogether. Section 2.4.1 goes into detail about the features that correspond to the Sentinel-2 bands, whereas Section 2.4.2 goes into detail about the features that refer to the multispectral indices.



Fig. 4. P-VG1 ground truth mask example.

2.4.1. Spectral bands

The first 13 features for a given pixel are the 13 Sentinel-2 bands. The remaining features are determined using these 13 bands in various combinations. Two satellites designated Sentinel-2A and Sentinel-2B constitute the Sentinel-2 mission. The orbit of each satellite lasts roughly 10 days. The revisit time is halved since both satellites are at their greatest separation from one another. Therefore, it takes about 5 days to get updated images of the same area. The wavelength, bandwidth, and spatial resolution of each band for the Sentinel-2A and Sentinel-2B satellites are displayed in [Table 1](#). Bands are interpolated using the Nearest Neighbor algorithm to the resolution of the band with the best spatial resolution, in this case 10 meters per pixel. The main visible (B02, B03 and B04) and near-infrared (B08) Sentinel-2 bands have a spatial resolution of 10 meters, while its red-edge (B05, B06 and B07), narrow near-infrared (Band 8A) and two shortwave infrared (B11 and B12) bands have a 20-meter spatial resolution. The coastal aerosol (B01), water vapor (B09), and cirrus (B10) bands have a spatial resolution of 60 meters. In the present study, bands B02, B03 and B04 are called "visible bands", and bands B05, B06, B07, B08, B08A, B11 and B12 are called "IR bands".

While each band provides unique information, it's crucial to acknowledge that not all bands are equally valuable for the specific task of manure detection. For example, bands B01 and B10 may not offer significant insights for the current study. The primary objective is to extract the maximum information from the available data. This study places emphasis on feature selection methods to identify the most relevant bands. For this reason, the approach follows a data-driven strategy, avoiding assumptions and maintaining scientific rigor by using all available bands.

2.4.2. Multispectral indices

The most prominent multispectral indices described in the literature for precision agriculture are reviewed to create as many significant features as possible. The formula for each multispectral index utilized in this investigation is described in the [supplementary material](#). The works ([Zhu et al., 2021](#); [Sishodia et al., 2020](#); [Shou et al., 2007](#); [Romanko, 2017](#); [Ma et al., 2010](#); [Fu et al., 2021](#); [Dodin et al., 2021](#); [Bagheri et al., 2013](#); [Sentinel-Hub, 2022](#)) describe all the details regarding the multispectral indices used in this study.

2.4.3. Feature selection methods

Using multiple multispectral indices usually improves accuracy over a single index as long as they are all correlated to the target ([Diaz-Gonzalez et al., 2022](#)). However, an excessive number of features can potentially have a negative impact on the accuracy of the classification model. This is because when more dimensions (features) are added, more data is needed to train the models. The increase in volume causes the data to become sparse, which is a phenomenon known as the "curse of dimensionality" ([Karanam, 2021](#)). Due to this reason and the

potential redundancy of multiple features, various approaches for feature selection are evaluated.

- Ward clustering algorithm ([Ward and Joe, 1963](#)): a method based on Euclidean distances to detect groups of features that are correlated. To help with visualization, a dendrogram is generated.
- Boruta method ([Kursa and Rudnicki, 2010](#)): utilizes a Random Forest classifier to choose features that are relevant to the output variable. The significance of a feature is calculated using the increase in model prediction error after varying the values of the feature. These modified features are called shadow features. A feature is relevant when shuffling its values increases the model error, as the model relied on the feature.
- Recursive Feature Elimination method ([Guyon et al., 2002](#)): uses the dataset to build a Decision Tree classifier employing the outcome variable. Then, evaluates the importance of the features and removes the least relevant feature. Once this step is done, the classifier is rebuilt and the process is repeated. This procedure is iterated until the desired number of features is reached.
- Principal Component Analysis method ([Wold et al., 1987](#)): simplifies the complexity of high-dimensional sample spaces while preserving their information by projecting the dataset into a smaller number of orthogonal dimensions, known as components, that retain the higher variance of the original data. This method belongs to the unsupervised learning as it does not use the outcome variable.
- Optimal Biomarker method (OB) ([Tong et al., 2018](#)): the first stage of the OB method involves selecting features based on their correlation and p-value. The number of features to keep is chosen manually at this stage. The second stage involves training a model and evaluating its performance using every possible combination of features. Any feature that does not improve performance at this stage is removed. To produce a statistically sound model, this second stage must be repeated N times.
- Optimal Biomarker Simplified method (OBS): the computational capacity of the OB method increases exponentially with the number of features. For this reason, it is not feasible to obtain a set of features much larger than 10. This simplified method uses only the first stage of the OB method to obtain the desired number of features greatly reducing its complexity.

In addition to the feature sets derived by the aforementioned methods, all Sentinel-2 bands, only the visible bands (B02, B03, and B04), only the IR bands (B05, B06, B07 B08, B8A, B09, B10, B11, and B12), all of the multispectral indices, as well as the subtraction between the features of both images, are individually proposed for evaluation in order to compare the significance of the different bands and multispectral indices.

Table 1

Central wavelength, bandwidth and spatial resolution of the Sentinel-2 bands. S2A is Sentinel-2A, and S2B is Sentinel-2B.

Band	Central Wavelength (nm)	Bandwidth (nm)	Spatial resolution (m)
	S2A - S2B	S2A - S2B	
B01 Coastal aerosol	442.7–442.2	21–21	60
B02 Blue	492.4–492.1	66–66	10
B03 Green	559.8–559.0	36–36	10
B04 Red	664.6–664.9	31–31	10
B05 Vegetation Red Edge (VRE)	704.1–703.8	15–16	20
B06 VRE	740.5–739.1	15–15	20
B07 VRE	782.8–779.7	20–20	20
B08 Near-Infrared (NIR)	832.8–832.9	106–106	10
B8A Narrow NIR	864.7–864.0	21–22	20
B09 Water vapour	945.1–943.2	20–21	60
B10 Short-Wave Infrared (SWIR) Cirrus	1,373.5–1,376.9	31–30	60
B11 SWIR	1,613.7–1,610.4	91–94	20
B12 SWIR	2,202.4–2,185.7	1.75–1.85	20

2.5. Performance evaluation metrics

Multiple metrics are computed to evaluate the performance of the models and facilitate the comparison of results obtained from different strategies (Fernandez-Moral et al., 2018). The notions of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) are the foundation of all metrics. TP stands for the quantity of correctly classified pixels. FP pixels are those that are categorized as the target class even though they belong to a different class. TN pixels are those that are legitimately not assigned to the target class. FN stands for pixels that were erroneously attributed to a different class.

One of the most popular metrics for expressing a model's performance in a single value is Accuracy. It is determined by dividing the overall number of pixels that were correctly classified by the overall number of pixels, as indicated in Eq. (1). If the amount of pixels in each class is too disproportionately distributed, this metric may be deceptive.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Recall is determined by dividing the correctly categorized pixels from the ground truth pixels that belong to the target class, as illustrated in Eq. (2).

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Precision is calculated by dividing the number of successfully classified pixels by the total number of detected pixels, as indicated in Eq. (3).

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

When Precision is low and Recall is high, it implies that there is a tendency to overclassify pixels from the target class. In contrast, when Recall is low and Precision is high, only pixels with a high degree of confidence will be classified as belonging to the target class. The F₁-Score is a commonly used metric in image segmentation, as it evaluates both Precision and Recall at the same time. It is determined by combining Precision and Recall, as shown in Eq. (4). The Mean F₁-Score is obtained by computing the Mean Precision and Mean Recall. Mean Precision and Mean Recall are calculated by averaging the Precision and Recall of both classes.

$$F_1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

3. Results and discussions

3.1. Dataset generation

The data gathering process involved visiting a total of 38 distinct plots. However, 8 of these plots were excluded from the study due to their small size, the absence of a precise date of manure spreading, or the presence of excessive cloud cover in the captured images during that period. To enable reproducibility, the entire dataset, including the analysis of each visited plot and the development of its ground truth, is publicly accessible ¹.

This region is known for its cool and damp climate due to the influence of the Atlantic Ocean. Winters are mild and rainy while summers are cool and cloudy. Rainfall is abundant resulting in a green landscape. For this dataset, there are no other plot types besides pasture lands and maize. This is because these are the most common in this region of northern Spain. The manure utilized in this study is a blend of liquid and solid cow manure, applied uniformly with the assistance of a manure

Table 2

Estimated date of manure application, coverage area and coordinates for every plot in the dataset.

Plot	Date (d/M)	Area (ha)	Coordinates (Long./Lat.)
P- BLD	26/05	0.89	-4.201, 43.397
P- BLLT1	16/05	2.12	-4.084, 43.430
P- BLLT2	26/05	0.33	-4.084, 43.431
P- CRDN	24/02	0.65	8.658, 45.859
P- CBRCS1	26/05	0.67	-4.200, 43.389
P- CBRCS2	26/05	0.64	-4.204, 43.387
P- CLGT	16/05	1.72	-4.109, 43.398
P- CLMBRS	26/05	0.43	-4.544, 43.380
P- CMNTR	16/05	0.26	-4.147, 43.400
P- DR	21/03	0.25	-4.142, 43.396
P- FNFR	16/05	1.01	-4.265, 43.388
P- LLT	03/05	0.96	-4.151, 43.400
P- LNDRS1	16/05	0.32	-4.251, 43.388
P- LNDRS2	16/05	0.54	-4.250, 43.388
P- LNDRS3	16/05	0.85	-4.249, 43.387
P- LNDRS4	16/05	0.91	-4.246, 43.387
P- MT	04/05	1.99	-4.153, 43.398
P- NMS	10/02	0.55	-4.149, 43.400
P- QNTLS2	16/05	0.85	-5.584, 43.545
P- SNTLLN	17/03	1.42	-4.117, 43.393
P- SNVCNT1	16/05	0.67	-4.404, 43.393
P- SNVCNT2	16/05	2.92	-4.400, 43.394
P- STBN	04/05	1.13	-4.136, 43.396
P- TGL2	16/05	1.23	-4.070, 43.427
P- TNNS1	26/05	1.95	-4.187, 43.399
P- TNNS2	26/05	1.58	-4.191, 43.398
P- VG1	09/04	1.22	-5.486, 43.469
P- VG2	13/04	0.49	-5.480, 43.469
P- VLDMR	07/02	1.75	-4.156, 43.405
P- VNS	23/04	1.66	-4.150, 43.404

spreader. The region of interest of the plots has an area of about 1,700 square meters, however, the plots are significantly smaller. This area includes all inspected plots as well as a few nearby manure-free areas. In this region, land use generally fall into one of the following categories:

- Tall grass. The color of this kind of pasture is a vibrant green.
- Freshly mown. The color of this kind of pasture is often yellowish green.
- Sowed/Plowed lands. The hue of this kind of pasture is brown. This type of land can be visually indistinguishable from manure application.
- Maize fields: Dark shadow lines and a dark green color characterize this kind of crop.
- Grazing lands. Similar to other varieties of pasture, this one is brown in hue. The livestock in the plot is responsible for its color and can be visually hard to distinguish from manure application.
- Woods. The color of this kind of pasture is dark green. Usually, it is composed of a large number of trees.
- Remnant habitat/Bushland. The hues of this kind of pasture range from a deeper green to brown. It is composed of a variety of wild grasses.
- Manure application: It can be used in recently mowed or seeded/plowed fields. The tint of the land shifts to a deeper hue. Due to the way it was dispersed, it is frequently found as circular marks.

In this work, 30 plots are investigated in total. Each plot's identifier, manure application date, and area is displayed in Table 2 for each plot.

The total dataset is composed of 225.94 hectares, of which 31.48

Table 3

Hectares of manure application and others in the dataset.

Class	Train	Test	Total
Manure application	22.28	9.20	31.48
Others	158.51	36.95	195.46

¹ <https://doi.org/10.17632/fbvfvf55kp.1>

Table 4
SIGPAC classes. Excluded classes are marked in gray.

Class	Description	Group
CF	Citrus-Fruit Association	Permanent crops
CS	Citrus-Fruit Peel Association	Permanent crops
CV	Citrus-Vineyard Association	Permanent crops
FF	Association Fruit Trees-Fruit Trees Of Peel	Permanent crops
OC	Olive-Citrus	Permanent crops
CI	Citrus	Permanent crops
FY	Fruit Trees	Permanent crops
FS	Dried Fruits	Permanent crops
FL	Nuts and Olives	Permanent crops
FV	Nuts and Vineyard	Permanent crops
OV	Olive grove	Permanent crops
OF	Olive grove - Fruit trees	Permanent crops
VI	Vineyard	Permanent crops
VF	Vineyard - Fruit Tree	Permanent crops
VO	Vineyard - Olive grove	Permanent crops
TA	Arable Land	Farmland
TH	Orchard	Farmland
IV	Greenhouses and crops under plastic	Farmland
PS	Pastureland	Pastures
PR	Shrub Grass	Pastures
PA	Pasture with Trees	Pastures
FO	Forest	Forest
AG	Watercourses and Water Surfaces	Non-agricultural area
ED	Buildings	Non-agricultural area
EP	Landscape element	Non-agricultural area
IM	Unproductive	Non-agricultural area
CA	Roads	Non-agricultural area
ZU	Urban Zone	Non-agricultural area
ZV	Censored Area	Missing data
ZC	Not included	Missing data

Table 5
Description of all feature sets for experimentation.

Feature Set	Description
A-13-S2	Sentinel-2 bands after manure application.
BA-26-S2	Sentinel-2 bands after and before manure application.
BA-6-RGB	Visible bands after and before manure application.
BA-16-IR	Infrared (IR) bands after and before manure application.
A-51-MI	All calculated multispectral indices after manure application.
BA-102-MI	All calculated multispectral indices after and before manure application.
A-64	All features after manure application.
BA-128	All features from after and before manure application.
A-7-W	7 manually selected features using Ward clustering after manure application.
BA-14-W	7 features from A-7-W and their equivalent from before manure application.
BA-95%-PCA	PCA components capturing 95% variability from all features.
BA-45-PCA	45 PCA components from all features (over 99% variability).
BA-90-RFE	90 features selected using Recursive Feature Elimination using all features.
BA-10-OB	10 selected features using Optimal Biomarker using all features.
BA-80-OBS	80 selected features using simplified Optimal Biomarker using all features.
BA-64-D	Difference of features between after manure application and before.
BA-8-EOMI	EOMI features (EOMI1, EOMI2, EOMI3, and EOMI4) extracted using all features.

hectares are classified as "Manure application" and 195.46 hectares are classified as "Others" which serves as a counterexample. A pixel is equal to 0.01 hectares (10 square meters). To avoid the training and test sets having adjacent pixels, plots rather than pixels are used to split the dataset. This lessens the chance of overfitting. 70% of the available plots, 21 of the 30 plots, are used for training, totaling 180.79 hectares (158.51 hectares for the "Others" class and 22.28 hectares for the target "Manure application" class). The remaining 30% of the plots, or 9 of the 30, are used to assess the models' performance, totaling 46.15 hectares

(9.20 hectares for the target class and 36.95 hectares for the "Others class"). This information is summarized in Table 3.

Pixels of no interest in terms of land usage are removed using the Geographic Information System of Agricultural Land (SIGPAC) database. Access to this database is free, and it is constantly updated. Only Spanish land use data is accessible, but this does not limit the use of this method because equivalent databases are available in other European Union nations.

There are 30 classes total in SIGPAC. 18 of these classes are deemed relevant, while 12 are not. Table 4 displays classes of no interest in gray. (see Table 5).

For the class "Others" the choice of counterexamples is made manually. Only pixels that can be confidently proven not to have been fertilized by on-site inspections of surrounding plots are chosen because not all feasible locations of each image could be validated. Additionally, only a portion of the image's potential pixels are chosen in order to prevent a severe class imbalance. The dataset contains 195.46 hectares in total of "Other" class counterexamples. 36.95 hectares are used for testing, and 158.51 ha are used for training.

3.2. Time series analysis

After manure application, grass grows back quickly, as shown in the example of P-VG1 from Fig. 5. This suggests that the visibility of the manure in the field greatly depends on the time the image is acquired after the manure application. When clouds hide the plot and the image is acquired long after the manure application, the discrepancy in values from the image before and the image after manure application could be diminished, greatly limiting the capability for detection.

It is important to highlight that the presence of freshly mown grass is a common occurrence just before the application of manure, as seen in P-SNVNT2 in Fig. 6. The first image shows the state of the plot two images preceding manure application, with a dark green color. The second image shows freshly mown grass, with a light green color, taken just before manure application. The third and final image shows the plot after manure application, with a dark brown color. This pattern can be found in many of the analyzed plots. However, this stage can only be seen for a few days. Some plots with the exact same pattern do not include this image because the satellite's revisit time is approximately 5 days. Therefore, in this study, the image immediately prior to manure application is used regardless of whether it is from recently cut grass or not.

3.2.1. Spectral band correlation with manured plots

In this section, the variability in intensity of the 13 Sentinel-2 bands during the time progression before and after the application of the manure is analyzed. Fig. 7 shows all band intensities for each of the plots. Intensity is calculated as the median of all the pixels in the plot of interest using the ground truth mask. A red area is added to the charts to indicate the period in which the plot was manured.

From Fig. 7 it can be determined that bands B06, B07, B08, B8A, and B09 are highly correlated to the manure in the plot. In all of the plots, their intensity value plummets from around 0.60 to about 0.40. The primary influencers affecting the intensity of the infrared (IR) bands appear to be the variations in heat and humidity associated with the application of manure. Then, as the grass starts to grow back and the manure is absorbed, the values of these bands start to recover.

3.2.2. Band visualization

Bands B06, B07, B08, B8A, and B09 produce the most notable difference in intensity when manure is applied. For this reason, to observe the most correlated bands they are displayed as images zoomed in the region of the plot (see Fig. 8) for the nearest days before and after manure application. This figure shows that a visual difference is appreciable.

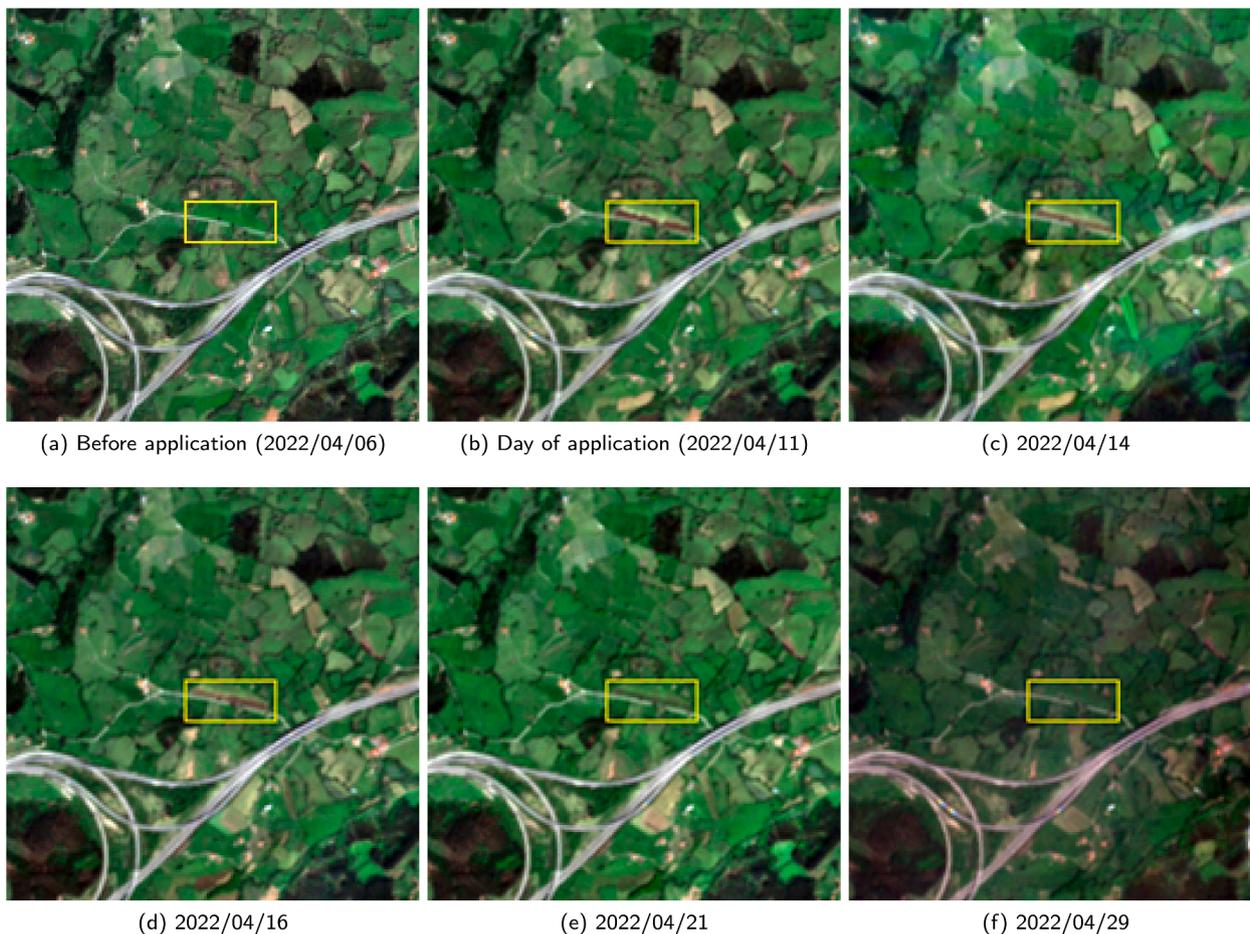


Fig. 5. Manure application time series of plot P-VG1.

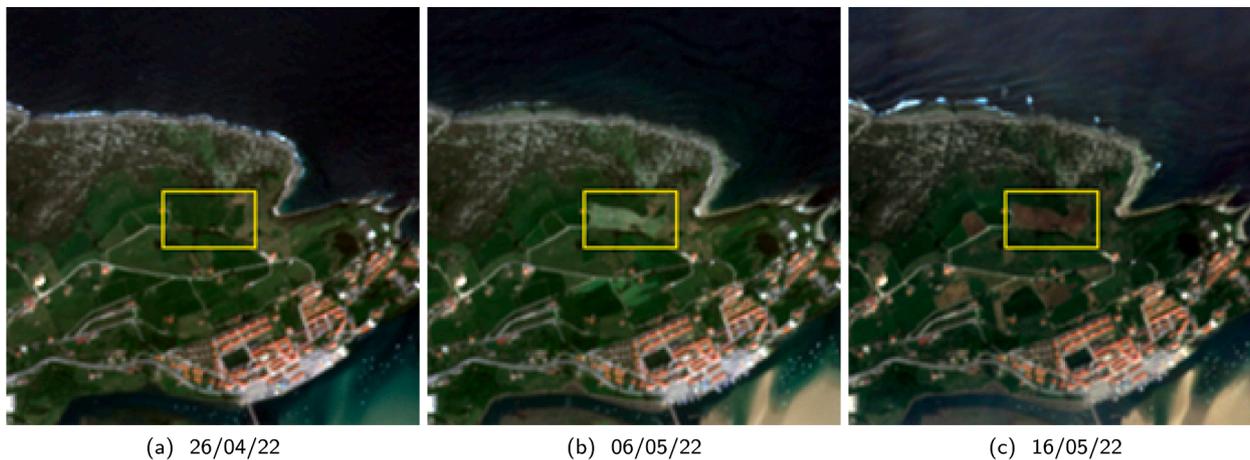


Fig. 6. Time series of the plot P-SNVCNT2. This example shows mown grass before manure application.

3.2.3. Multispectral indices correlation

In this section, the results of the 51 multispectral indices during the time progression before and after the application of the manure is analyzed. There are too many values to be represented in a line graph, thus, to represent their intensity a heatmap with all 51 multispectral indices for each plot is shown in Fig. 9. Values are normalized per row (the range [min, max] is mapped to [0, 1]). In the X axis, the first image after manure application is identified as image 0. Images before or after manure application are labeled as negative or positive, respectively. A vertical dotted red line is drawn at the start of the manure application

day to facilitate visualization. In Fig. 9, the intensity of most multispectral indices is close to 0 the day of manure application, and then, as time passes, the values increase. Furthermore, the rest of the indices appear to be reversed (inversely correlated), with values close to 0 before manure application, and close to 1 when manure is applied. Only a few of the multispectral indices seem to be uncorrelated to manure application. Therefore, the considered indices seem to provide valuable information about the existence manure in the plots.

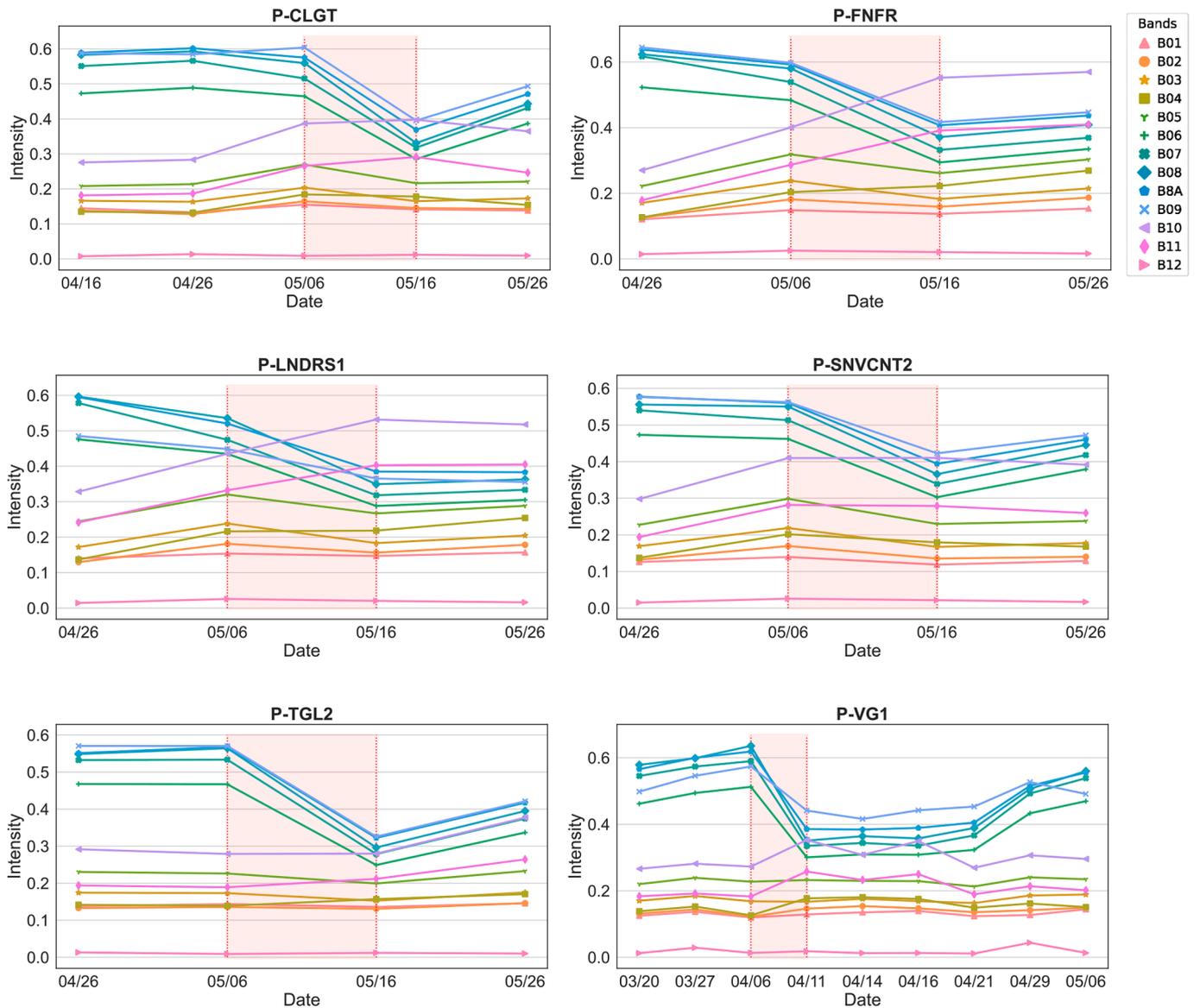


Fig. 7. Intensity time series of the bands for P-CLGT, P-FNFR, P-LNDRS1, P-SNVCNT2, P-TGL2, and P-VG1.

3.3. Feature correlation study

In this section, a study about the correlation of all the features is performed to analyze feature redundancy. Features consist of the 13 Sentinel-2 bands, as well as the 51 multispectral indices. For each pixel, there are a number of samples corresponding to the acquired images before and after manure application. To study the correlation between all features, the pairwise Pearson's linearity correlation coefficient is calculated for all possible combination of features (see Fig. 10). This correlation is calculated with Eq. (5).

$$r = \frac{n \sum x_i y_i - \sum x_i - \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} - \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \quad (5)$$

Fig. 10 shows that the majority of features are highly correlated with at least a few other features. It is reasonable to expect that closely situated bands exhibit strong correlations, as do similar indices. Formulas for each multispectral index can be found in the supplementary material. TSAVI is the only feature that is not correlated to any other feature.

The TSAVI index, expressed as $\frac{(0.421 * (B08 - 0.421 * B04 - 0.824))}{(B04 + 0.421 * (B08 - 0.824) + 0.114 * (1 + 0.421)^2)}$, utilizes band B04 (Red with 10 meters of spatial resolution) and band

B08 (NIR with 10 meters of spatial resolution), introduces a factor which typically set to 0.421. This factor introduces a significant distinction compared to other indices.

The correlation between the first 13 bands is studied next. The first five bands are correlated with each other (B01, B02, B03, B04, and B05). Bands B06, B07, B08, B8A, and B09 are heavily correlated. Finally, B10 and B11 are also correlated. B12 is the one with weaker correlation to the rest.

A correlation greater than 0.80 in absolute value indicates data redundancy. To determine the uniqueness of the features, the number of features with a correlation equal to or greater than 0.80 is counted. In this way, the percentage of the total number of features in which the index is highly correlated can be obtained. This means that a feature with a high ratio, for example with a value of 0.6, is highly correlated to 60% of all other features. From this information, a bar graph is shown in Fig. 11. The following features are highly correlated to less than 10% of all features: B01, B02, B03, B05, B10, B12, EOMI2, EOMI4, BNR2, CI2, CI3, NBRI, NDR2, NDR3, TSAVI, CARI1, CARI2, CVI. However, they could still be highly correlated with each other.

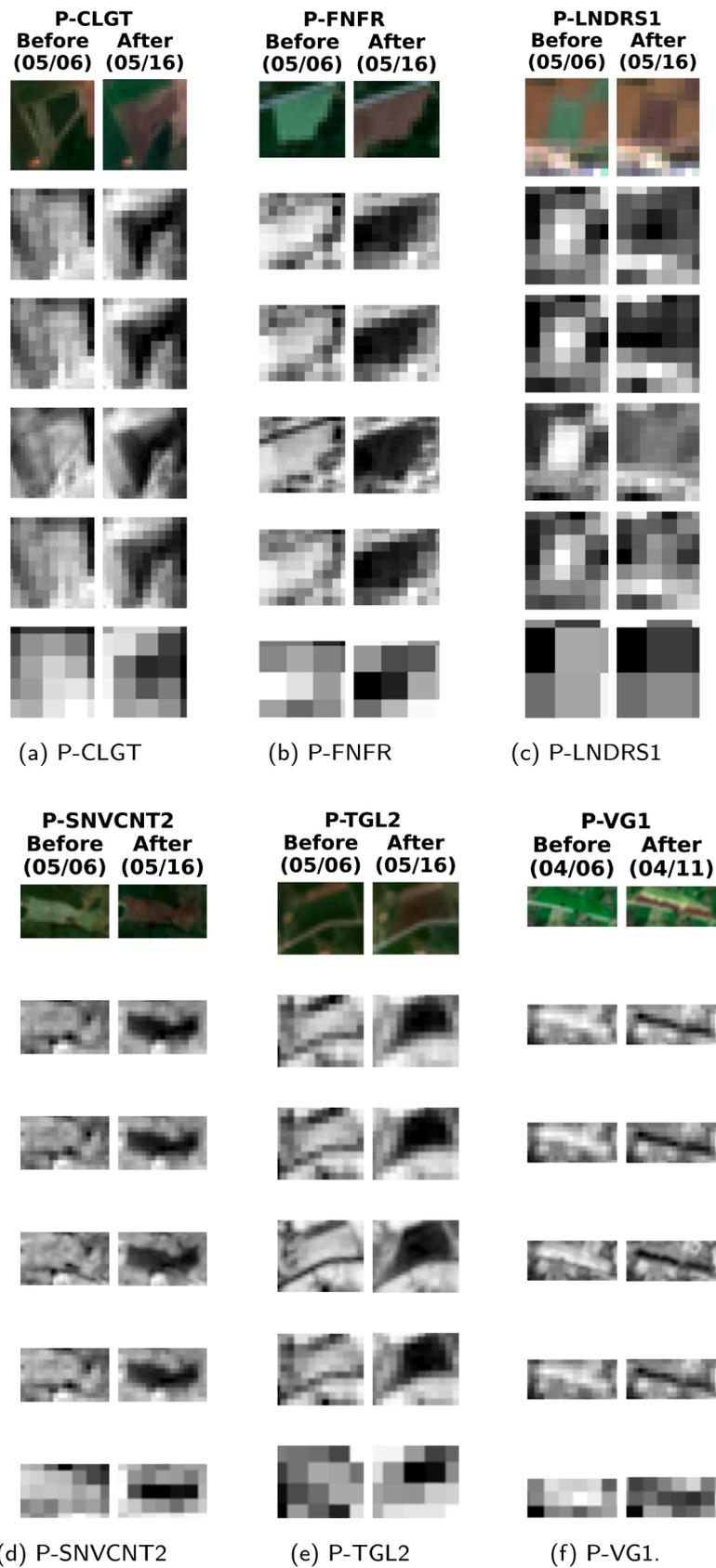


Fig. 8. Visual differences before and after manure application in bands RGB, B06, B07, B08, B8A, and B09.

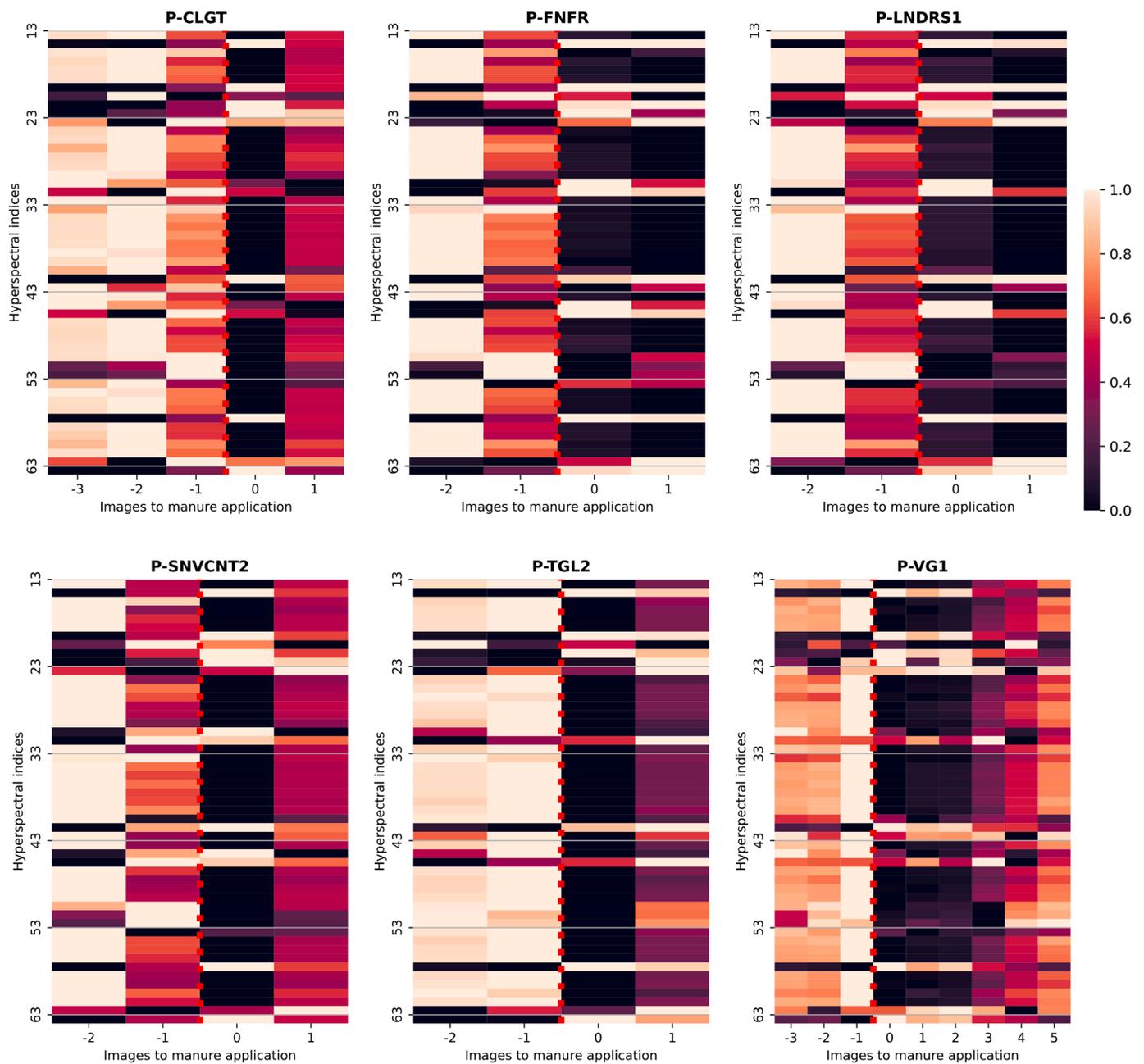


Fig. 9. Intensity for all multispectral indices before and after manure application.

3.4. Results of the feature selection methods

In this section different feature selection methods are applied to the dataset. The objective is to select a set of features without redundant or irrelevant elements.

- **Ward:** this method is used to group the features in clusters. The dendrogram from Fig. 12 shows that the 64 features can be reduced to 7 groups. A feature from each group is selected. The 7 manually selected features and the reason for their selection are shown below:
 1. NBRI: More commonly used in agriculture than the rest.
 2. B04: Has the most spatial resolution and does not require further calculations.
 3. B03: Has the most spatial resolution and does not require further calculations.
 4. NDVI: More commonly used in agriculture than the rest.
 5. EOMI4: Specific for exogenous organic manure identification.

6. TSAVI: Single feature in cluster.
7. EOMI2: Uses B11 and B04 bands (B04 has more spatial resolution).

Fig. 13 shows that there is little to no correlation between the selected features (NBRI, B04, B03, NDVI, EOMI4, TSAVI, and EOMI2). This means that there are no redundant features as they show different information. This does not mean that they are more suitable for manure detection (relevant), but that they are not redundant.

- **Boruta:** this method considers that all features are important and does not discard any of them. This method does not reduce the number of features to be used therefore, the results are obviously the same as when using all the data (A-64 and BA-128).
- **Recursive Feature Elimination (RFE):** this method is used to evaluate the selection of 5, 10, 20, 30, 40, 60, 80, 90, 100, 110, and 120 features, out of a total of 128 features (using the 64 from the image before the manure application and the 64 from the image after

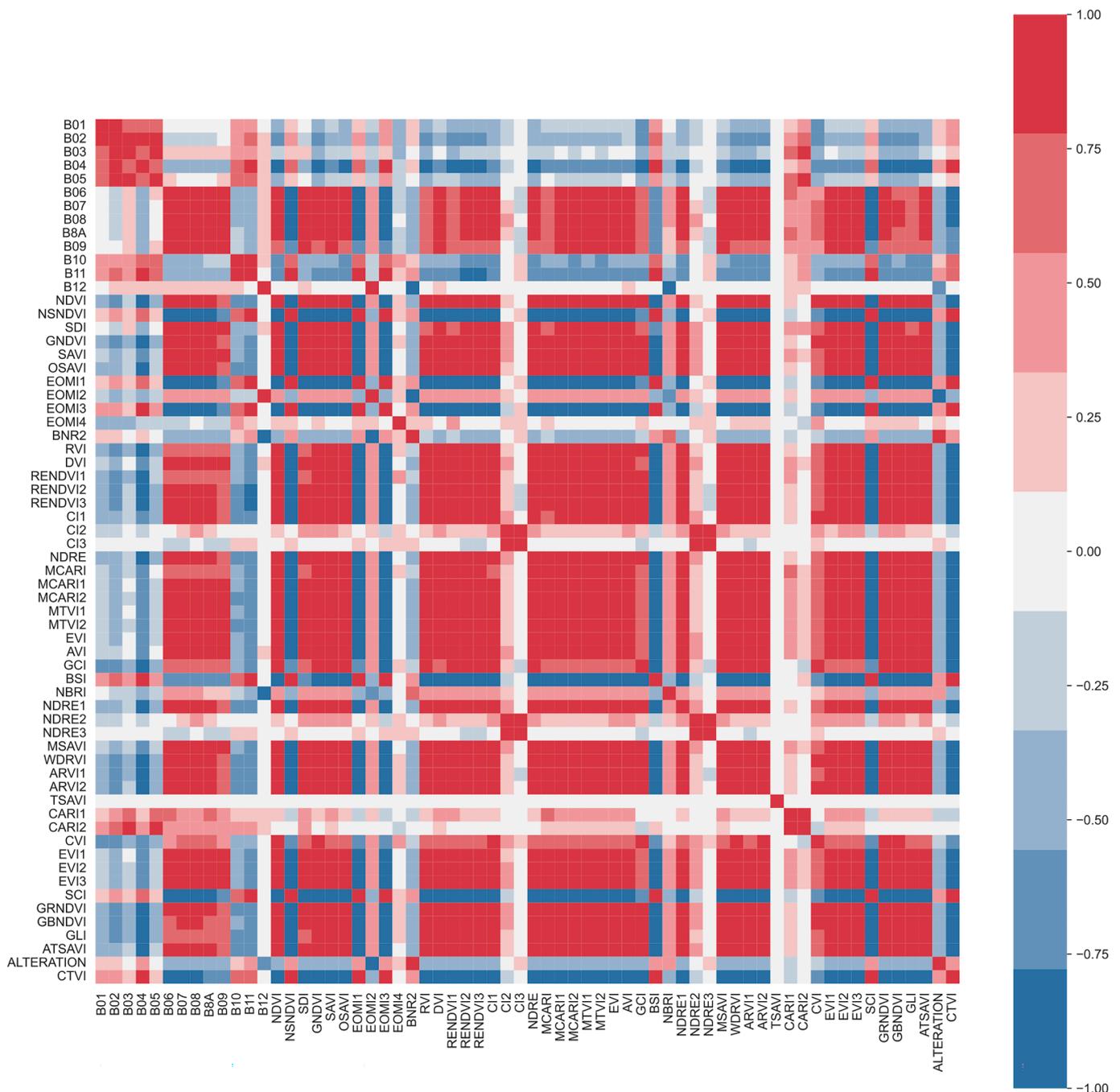


Fig. 10. Correlation study for all multispectral indices.

manure application as separate features). It is found that the best experiment uses 90 features (BA-90-RFE). Increasing the number of features causes the F_1 -Score of the model to decrease (see Fig. 14). Only the best experiment is documented.

- Principal Component Analysis:** this method is used for dimensionality reduction. All 128 features are used (64 from the image from before, and 64 from the image after manure application). The experiment which uses PCA as a detection model with a 95% variance is called BA-95%-PCA. This experiment uses 2 transformed features to achieve this variance. Another experiment called BA-45%-PCA is documented. This experiment uses 45 transformed features resulting in 99.999994% variance. The number of transformed features was selected as the one that provides the best results out of multiple experiments using 20, 30, 35, 40, 45 and 50 transformed features.

- Optimal Biomarker (OB):** this method is used to find the best features from the set of 64 features of the images after manure application. From these features two experiments are carried out: A-10-OB, which uses the features with only the image after manure application; and A-20-OB, which utilizes these features for the image from before and after manure application. The features found are: B02, B05, B10, B11, B12, EOMI4, BNR2, NBRI, NDRE2, NDRE3. In addition, the 10 best features are found considering the two days as separate features (before and after manure application). A total of 10 features out of 128 were obtained. The experiment that uses these features is called BA-10-OB. The features found are: B12 (before), MCARI (before), NBRI (before), B12 (after), CTVI (after), NBRI (after), GLI (after), B04 (after), B11 (after), B10 (after).
- Optimal Biomarker Simplified (OBS):** this method is used to evaluate the selection of 5, 10, 20, 30, 40, 60, 80, 90, 100, 110, and

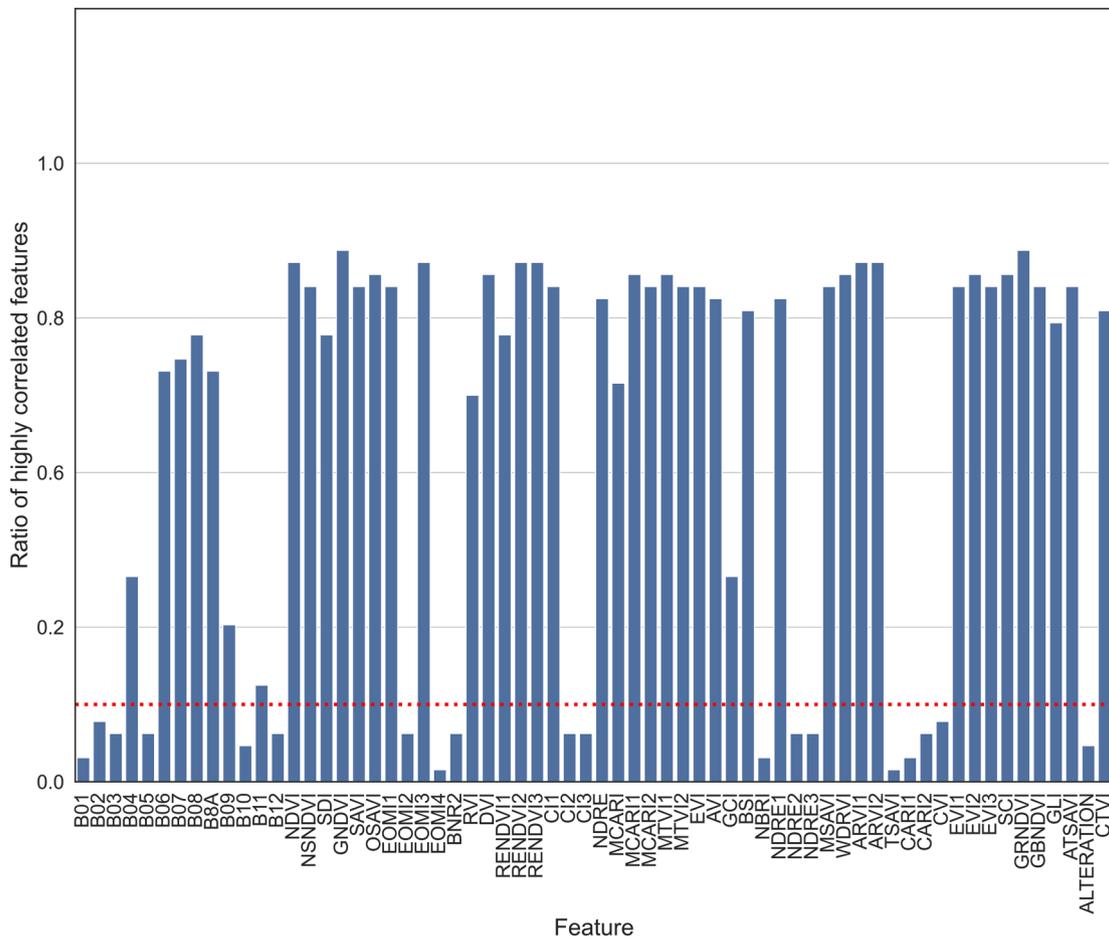


Fig. 11. Percentage of features each feature is correlated to. A red dotted line is drawn at 10%.

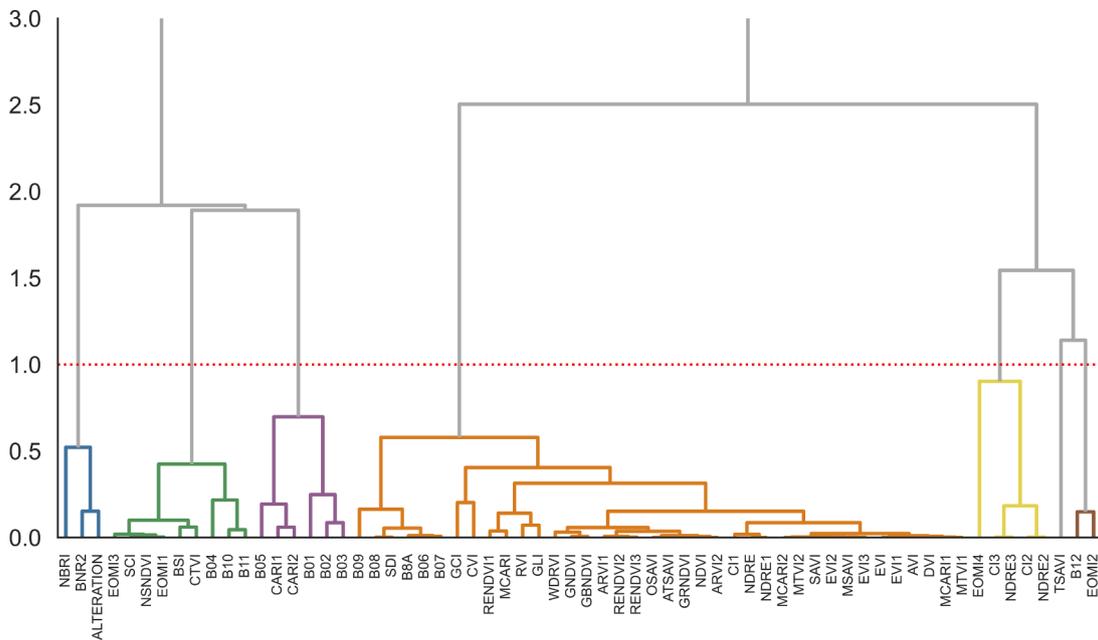


Fig. 12. Clustering of every feature. Seven groups are formed.

120 features, out of a total of 128 features (using the 64 from the image before the manure application and the 64 from the image after manure application as separate features). The best experiment uses 80 features (BA-80-OBS) as can be seen in Fig. 14. When going from

80 to 90 features its F_1 -Score decreases significantly. Even if the number of features continues to increase, the model remains constant.

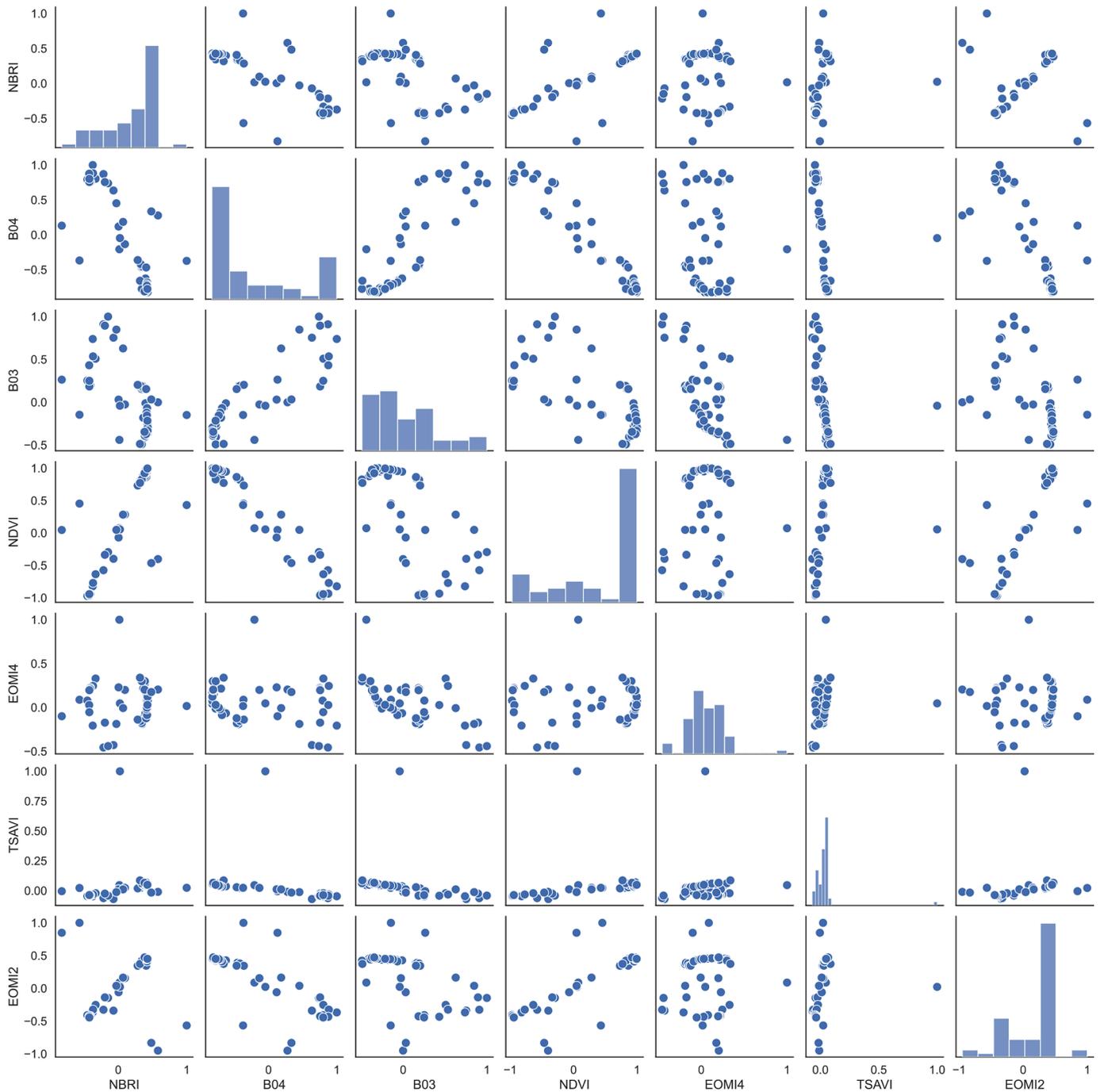


Fig. 13. Detailed correlation graph for the selected features.

3.5. Summary of experiments

The experiments involved the generation of multiple feature sets through various feature selection methods. Each selected classification method was trained and evaluated using these feature sets. This approach enabled the comparison of different feature sets and facilitated the determination of the optimal set. To uniquely identify each feature set, they were assigned identifiers such as "A-13-S2".

The identifiers consist of a combination of letters and numbers. The first part of the identifier represents the source or type of data used in the feature set. For example, "A" represents "After manure application" or "BA" represents "Before and After". The number in the middle indicates the number of features included in the set. For instance, "13" signifies 13 features. The last part of the identifier denotes additional information

about the feature set, such as the type of data or method used. For example, "S2" represents Sentinel-2 bands, "RGB" represents the visible bands, "IR" represents infrared bands, "MI" represents multispectral indices, "PCA" represents Principal Component Analysis, "RFE" represents Recursive Feature Elimination, "OB" represents Optimal Biomarker, "OBS" represents simplified Optimal Biomarker, "D" represents subtraction, and "EOMI" represents specific EOMI features. By combining these elements, the identifiers provide a concise representation of the feature set's characteristics and composition.

The best model for each feature set is displayed in Table 6. The train set is used to train all models, while the test set is used to compute all metrics.

This study provides evidence that the application of manure in agricultural fields can be accurately detected using satellite remote

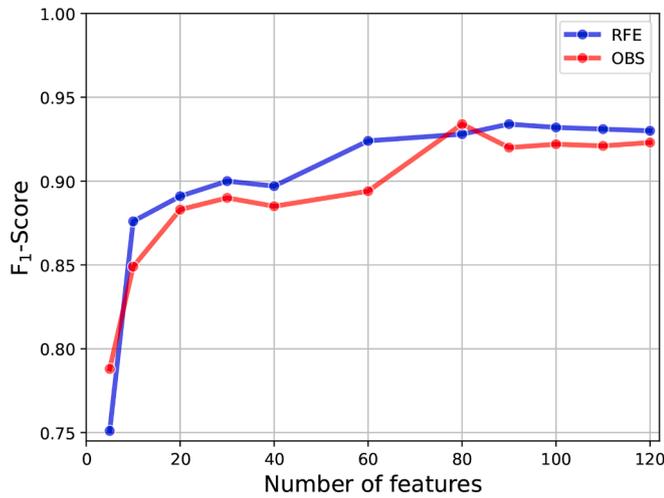


Fig. 14. F1-Score per number of features in RFE and OBS.

sensing techniques, yielding excellent results. The feature sets BA-102-MI, BA-90-RFE, and BA-80-OB show very similar results. The best models have a Mean F1-Score of 93.4%. However, these results were achieved using a dataset of a particular type of crop. With a larger and more varied dataset conclusions could change.

It was observed that using the features from the images taken prior to and following the application of manure produced superior models than those experiments that used features from only one image. For instance, while retaining the same Mean Precision, the Mean Recall of BA-128 is nearly 8% superior to that of A-64. This enhancement is true for all experiments of type “BA” against “A.”

Table 6

Results for each feature set experiment. Arranged in ascending order by F1-Score.

Feature set	Best classifier	Overall accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
BA-95%-PCA	Nearest Neighbor Classifiers	80.2	50.2	90.1	64.5
BA-8-EOMI	Decision Tree	88.9	77.6	85.4	81.3
A-7-W	Naïve Bayes Classifiers	88.8	82.4	82.5	82.4
BA-6-RGB	Neural Network Classifiers	90.6	82.3	86.9	84.6
BA-10-OB	Ensemble Classifiers	90.9	79.7	90.5	84.7
BA-14-W	Support Vector Machines	90.8	81.4	88.2	84.7
A-10-OB	Discriminant Analysis	90.5	84.3	85.5	84.9
A-13-S2	Support Vector Machines	91.5	79.9	92.9	85.9
BA-20-OB	Support Vector Machines	92.0	83.8	90.1	86.8
A-64	Ensemble Classifiers	92.7	82.2	94.8	88.1
A-51-MI	Support Vector Machines	92.8	83.9	93.0	88.2
BA-16-IR	Support Vector Machines	93.1	86.3	91.2	88.7
BA-26-S2	Discriminant Analysis	93.5	84.4	95.1	89.4
BA-64-D	Discriminant Analysis	94.0	86.8	94.1	90.3
BA-128	Discriminant Analysis	95.1	90.7	93.5	92.1
BA-45-PCA	Discriminant Analysis	95.2	89.3	95.4	92.3
BA-102-MI	Discriminant Analysis	95.9	91.6	95.2	93.4
BA-80-OBS	Discriminant Analysis	95.9	91.3	95.6	93.4
BA-90-RFE	Discriminant Analysis	95.9	91.0	96.0	93.4

Table 7

Results for the BA-102-MI experiment per classification method. Arranged in ascending order by F1-Score.

Classifier	Overall accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
Kernels Approximation Classifiers	85.2	67.1	81.4	73.6
Nearest Neighbor Classifiers	87.1	77.1	80.5	78.7
Decision Tree	88.1	76.9	83.3	80.0
Naïve Bayes Classifiers	85.1	87.9	77.9	82.6
Logistic Regression Classifiers	90.4	91.1	83.8	87.3
Support Vector Machines	93.3	83.6	95.5	89.1
Neural Network Classifiers	94.6	90.6	92.3	91.4
Ensemble Classifiers	94.9	88.4	95.3	91.7
Discriminant Analysis	95.9	91.6	95.2	93.4

In general, the more features the better results. The Ward clustering feature selection method (A-7-W and BA-14-W) shows results about 5–7% lower than when using all features. However, the Recursive Feature Elimination method (BA-90-RFE), the Optimal Biomarker Simplified method (BA-80-OBS), and using every multispectral index without the Sentinel-2 bands (BA-102-MI), all achieve better results than using all features (BA-128). Using more features than necessary can cause the models to be harder to train and produce worse detections. This problem is known as the “curse of dimensionality”.

The results obtained from BA-RGB indicate that good performance can be achieved by using only the visible bands of both images (before and after manure application), possibly due to their spatial resolution of 10 meters per pixel. However, the results obtained from BA-IR outperformed those of BA-RGB by approximately 4%, even though the IR bands have a lower spatial resolution of 20 meters per pixel. These findings suggest that the data from the IR wavelengths play a critical role in accurately detecting manure application. If all the IR bands have 10-meter spatial resolution as the visible B02, B03 and B04 bands, it is expected that even better results would be achieved. Additionally, the wavelength range of the IR bands is between 700 nm and 2200 nm. Extending this range to collect more information, by introducing new bands with increased wavelengths, may further improve the results.

BA-64-D performs better than A-64, this is because it contains more information since its features include data from both images. In this case BA-128 is still a superior model. This means that the subtraction of both features does not keep enough information.

BA-95%-PCA obtained much lower results than the rest of the feature sets. This is due to the fact that it does not manage to keep much information in its 2 components. In this sense BA-45-PCA, which contains 45 components instead of 2, obtains much better results. However, the full set of 128 features is used to calculate these 45 components. When comparing its results with BA-128, a slight improvement is observed,

Table 8

Class-specific results for Discriminant Analysis of the BA-102-MI feature set.

Class	Recall	Precision	F ₁ -Score (%)
Manure application	84.5	94.2	89.1
Others	98.7	96.2	97.4

although it does not justify the use of PCA because other feature sets such as BA-102-MI outperform this method.

A-10-OB achieves similar results to other methods with a comparable number of features such as A-13-S2 or A-7-W. The same occurs with BA-

20-OB, which obtains better results than BA-14-W but worse than BA-26-S2 or BA-16-IR. It is interesting to note that BA-10-OB, which obtains 10 features out of 128 from both images separately, obtains similar results as A-10-OB which obtains 10 features out of 64 from a single image. However, BA-80-OBS is the method that obtains the best models with the least number of features.

BA-8-EOMI is used to compare the multispectral indices from (Dodin et al., 2021), since these are the closest indices in the literature to manure detection. The results of this feature set are similar to experiments with a comparable number of features. This demonstrates that the

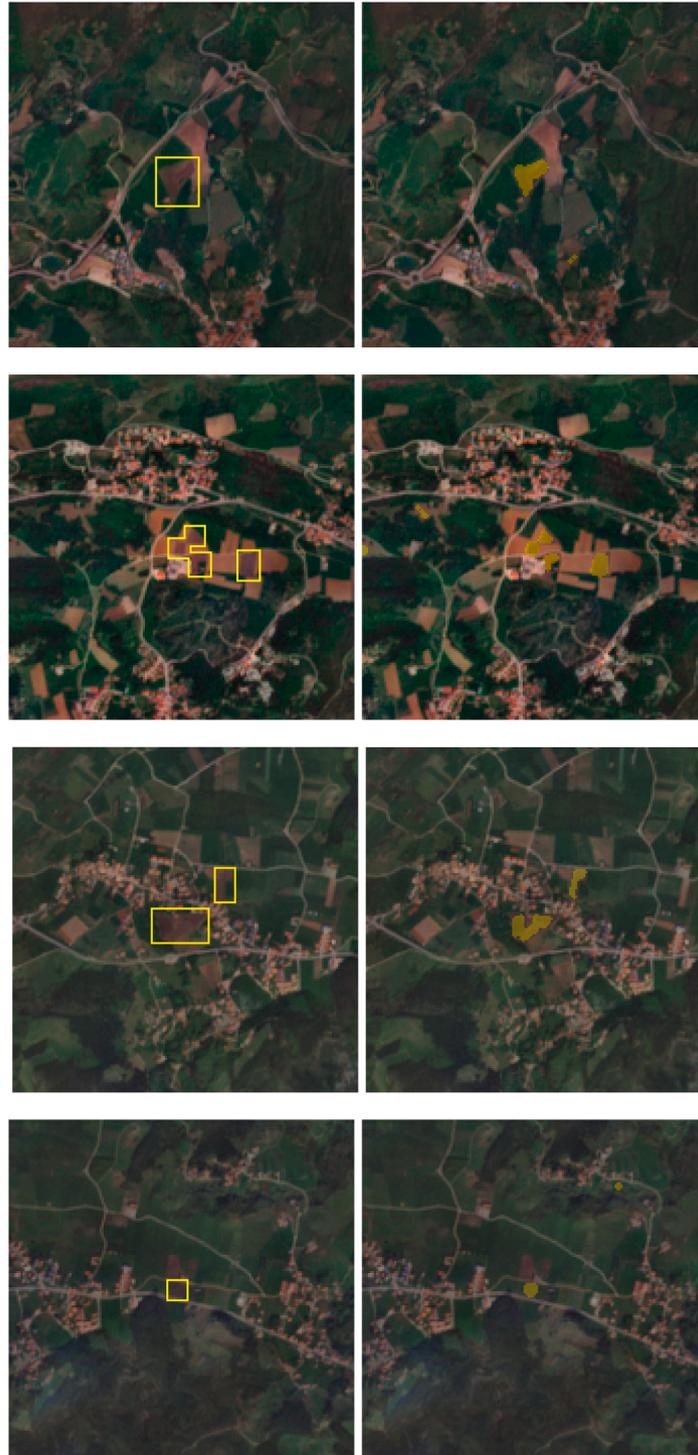


Fig. 15. Model detections in the test set.

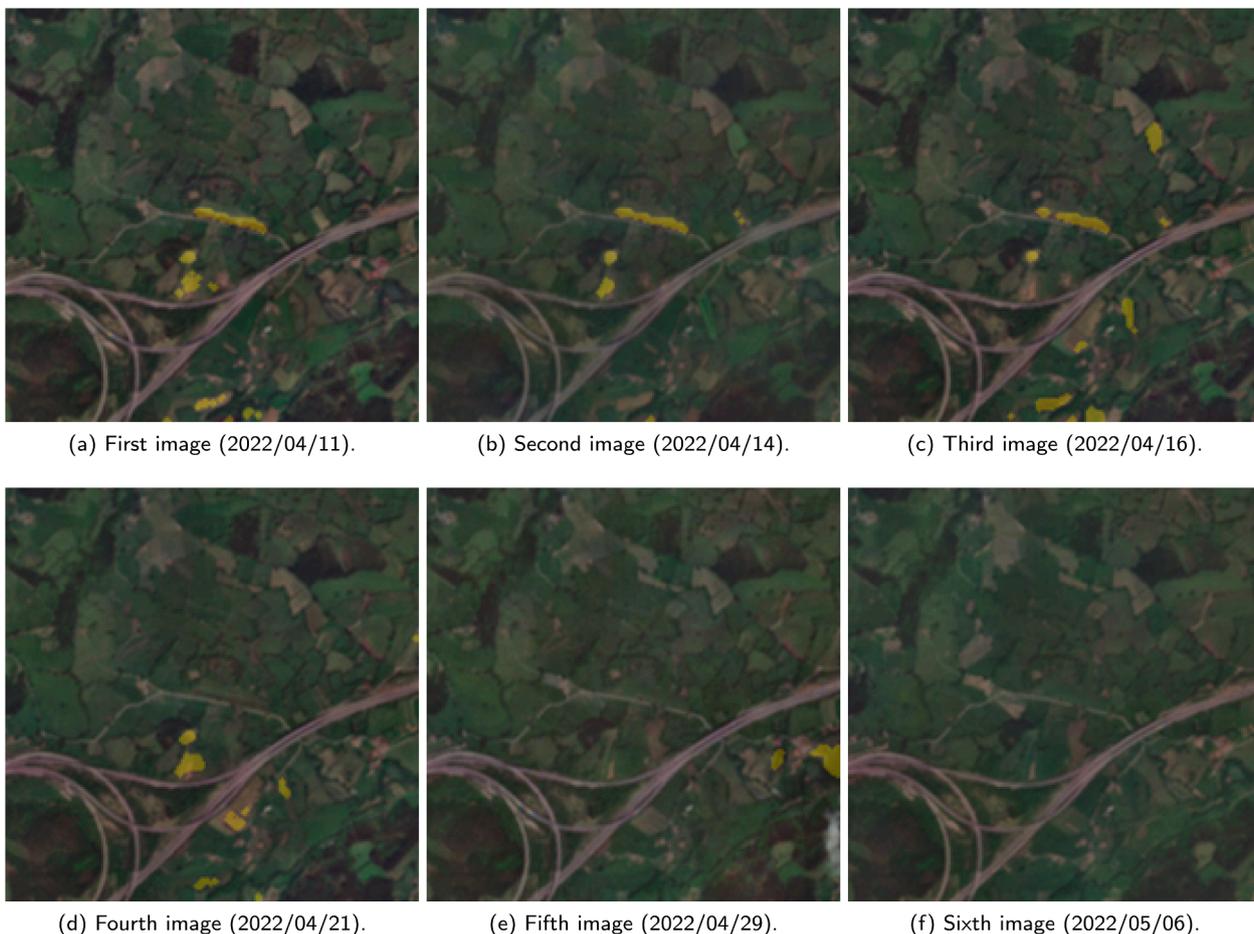


Fig. 16. Classification for different days after manure application of plot P-VG1.

use of multiple multispectral indices is superior to any single index in the current literature.

According to the results, Discriminant Analysis emerges as the classification method that consistently achieves the highest accuracy, particularly when a large number of features are utilized. On the other hand, Support Vector Machines generally perform better when the feature set is relatively smaller.

3.5.1. Experiment with the best result

The BA-102-MI set of features was selected as the one with the best results. Although its results are similar to those obtained by BA-90-RFE and BA-80-OBS, its metrics are slightly more balanced. In this section, the results of this best experiment are shown in detail. Table 7 shows the results for each classification method. The findings indicate that the Discriminant Analysis classification method outperforms all others, with a Mean F₁-Score exceeding the second best approach, Ensemble Classifiers, by nearly 2%.



Fig. 17. Monitoring simulation of plot P-VG1.

Table 8 shows the details of the best classification method. This table provides a breakdown of Precision, Recall, and F₁-Score for both the "Others" class and the "Manure application" class. It is worth noting that due to the higher number of samples in the "Others" class, the classes are unbalanced. Therefore, it is more appropriate to focus on the F₁-Score for the "Manure application" class, as it considers both Precision and Recall. The F₁-Score for the "Manure application" class is 89.1%, which implies that this model is proficient in identifying fields where manure has been applied on a per pixel basis.

The visual representation of manure detection in the test set is shown in Fig. 15. The left column displays the ground truth locations, while the right column represents the corresponding detection masks. The first row corresponds to the P-CLGT plot, followed by the second row containing P-LNDRS1, P-LNDRS2, P-LNDRS3, and P-LNDRS4. The third row showcases P-LLT and P-MT, and finally, the fourth row displays the detections for P-DR. The newly fertilized plots are almost perfectly detected, with few to no pixels that were incorrectly categorised. Only a little area outside the ground truth is wrongly categorized as the "Manure application" class. This occurs in most of the images. Outside of the groundtruth, there are some areas that are unknown and cannot be used to assess the effectiveness of the detection. This is merely a representation of the detection of the test plot images.

3.6. Time series experimentation

3.6.1. Maximum number of days after manure application

To study the maximum number of days after manure application in which the plots can still be detected, a visual detection of plot P-VG1 at different dates is used. To achieve the greatest results, the best experiment model (BA-102-C) is used. In all detections the same image from before manure application is maintained. Fig. 16 shows the progression of the detection as time passes. The second and third images after manure application are still detected successfully, however, the rest of the images are not detected. This shows that after approximately 10 days after manure application, this method is not able to detect manure in the plots.

To check the robustness of the approach, the best model (BA-102-C) is evaluated using an image after the one used during the test. In this way, the image before the application of the manure is kept the same, and the image after application is exchanged for the next one. Thus, an additional 5 to 10 days are usually added to the images showing the application of manure.

3.6.2. Monitoring simulation

In this section, the P-VG1 plot is used to study the behaviour of the classification model as the time sequence progresses. The idea of this experiment is that an image and its immediate subsequent image are used to find recently manured fields. The objective is to simulate a real monitoring scenario. It is expected that it will only be able to detect manure correctly at the time when the first image is from before manure application and the subsequent one is from just after manure application.

Fig. 17 shows possible image pairs in the time sequence. It can be seen that the plot is only completely found when the before and after images are right before and after manure application. The rest of the cases do not detect manure in the plot, or only find a small part of it.

4. Conclusion

This research assesses the effectiveness of widely used multispectral indices in precision agriculture and their connection to manure application, employing diverse feature selection methods. The aim is to identify potential legal infringements and mitigate leaching contamination resulting from manure application during periods of intense

rainfall. This study fills a research gap, as limited literature exists on this subject. The findings demonstrate the remarkable accuracy achieved in detecting freshly manured fields by utilizing a combination of multiple multispectral indices.

Sentinel-2 is the best public option for multispectral satellite imagery given its revisit time and spatial resolution. However, its spatial resolution is not great, where each pixel is either 10, 20 or 60 meters. To improve the analysis and detection of manured lands, image time series are used. The more relevant information is provided, the better the resulting model. Collecting multiple images in a sequence can be a difficult task because clouds can cover the plots. It has been found that using an image before manure application and another image immediately after can produce reliable models. Thus, minimizing the number of images needed to produce good results. The maximum time to acquire images after manure application has been found to be about 10 days. After 10 days, the plots tend to be visually indistinguishable as the rain begins to fall, the grass starts to grow, and the manure begins to decompose. Models are not able to detect residual manure. This time interval was calculated using images from spring, when grass grows most rapidly.

A total of 31.48 hectares of freshly manured pastureland plots were collected to train and test the models. Plots are located in northern Spain and were manually validated by on-site investigations. The features extracted from these images consist of the Sentinel-2 bands and the calculated multispectral indices. After analysis, it is observed that most of the features are observed to be highly correlated with the application of manure. To determine whether the features are redundant or add sufficient information, a feature selection study is performed. The experimentation carried out in this paper shows that a higher number of features is beneficial for the models up to a certain point. The Optimal Biomarker Simplified feature selection method shows that the minimum number of features for maximum model accuracy is 80. Using 90 features from the Recursive Feature Elimination method, as well as all the 102 multispectral indices without the 26 bands provides similar results. When more features are added, the model starts to decrease in accuracy. This could be caused by the known problem of the "curse of dimensionality". In the same manner, when less features are used, the models do not have enough data to achieve high accuracy. Infrared bands are able to differentiate manure application in fields with more accuracy than the visible bands even though they have half the resolution. However, bands alone cannot achieve accuracy levels as high as multispectral indices. Multispectral indices are essential to train a good model, but it has been found that no particular multispectral index to detect manure exists, multiple indices must be used for high accuracy. Thus, the optimal number of features is found to be between 80 and 102 out of the total 128 proposed.

CRedit authorship contribution statement

Oscar D. Pedrayes: Investigation, Data curation, Writing - original draft, Conceptualization, Methodology, Software. **Rubén Usamentiaga:** Supervision, Resources, Data curation, Conceptualization, Methodology. **Yanni Trichakis:** Resources, Validation. **Faycal Bouraoui:** Resources, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Dataset is available in: <https://doi.org/10.17632/fbvfvf55kp.1>

Acknowledgments

This work has been supported by the project PID2021-124383OB-I00 of the Spanish National Plan for Research, Development, and Innovation, as well as by the Severo Ochoa grant with reference PA-22-BP21-130.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.ecolind.2024.111550>.

References

- Aizerman, A., 1964. Theoretical foundations of the potential function method in pattern recognition learning. *Automation Remote Control* 25, 821–837.
- Alvarez-Vanhard, Emilien, Thomas Corpetti, and Thomas Houet (2021). "UAV & satellite synergies for optical remote sensing applications: A literature review". In: *Science of remote sensing* 3, p. 100019.
- Analytics, EOS Data (May 2022). Satellites Vs. Drones For Agri-Business: Use And Comparison. <https://eos.com/blog/drones-vs-satellites>. Accessed: 2023-03-30.
- Bagheri, Nikrooz, Ahmadi, Hojjat, Alavipanah, Seyed Kazem, et al., 2013. Multispectral remote sensing for site-specific nitrogen fertilizer management. *Pesquisa Agropecuária Brasileira* 48, 1394–1401.
- Berrar, Daniel (2018). "Bayes' theorem and naive Bayes classifier". In: *Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics* 403, p. 412.
- Bouma, J., 2016. The importance of validated ecological indicators for manure regulations in the Netherlands. *Ecol. Indicators* 66, 301–305.
- Brugger, Daniel and Wilhelm M Windisch (2015). "Environmental responsibilities of livestock feeding using trace mineral supplements". In: *Animal Nutrition* 1.3, pp. 113–118.
- Colwell, Robert N (1966). "Uses and limitations of multispectral remote sensing". In: *Council of the European Union* (1991). "Council Directive 91/676/EEC of 12 December 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources". In: *Official Journal* 375.
- Colwell, Robert N (2000). "Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy". In: *Official Journal*.
- Cover, Thomas and Peter Hart (1967). "Nearest neighbor pattern classification". In: *IEEE transactions on information theory* 13.1, pp. 21–27.
- Curran, Paul (1980). "Multispectral remote sensing of vegetation amount". In: *Progress in physical geography* 4.3, pp. 315–341.
- Diaz-Gonzalez, Freddy A., Vuelvas, Jose, Correa, Carlos A., et al., 2022. Machine learning and remote sensing techniques applied to estimate soil indicators—review. *Ecol. Ind.* 135, 108517.
- Dietterich, Thomas G (2000). "Ensemble methods in machine learning". In: *International workshop on multiple classifier systems*. Springer, pp. 1–15.
- Dodin, Maxence, Smith, Hunter D., Levasseur, Florent, et al., 2021. Potential of Sentinel-2 Satellite Images for Monitoring Green Waste Compost and Manure Amendments in Temperate Cropland. *Remote Sensing* 13 (9), 1616.
- Earth, Supervision (June 2019). Satellite VS Drone Imagery: Knowing the Difference and Effectiveness of Supervision Earth's Technology. <https://medium.com/supervisionearth/satellite-vs-drone-imagery-knowing-the-difference-and-effectiveness-of-fahmsupervision-earth-90e98b78777c>. Accessed: 2023-03-30.
- Fahmy, Fatma (2022). Pollution erodes fish stocks and livelihoods in Egyptian lake. <https://www.reuters.com/world/africa/pollution-erodes-fish-stocks-livelihoods-egyptian-lake-2022-09-01/>. Accessed: 2022-09-07.
- Fernandez-Moral, Eduardo, Martins, Renato, Wolf, Denis, et al., 2018. A new metric for evaluating semantic segmentation: leveraging global and contour accuracy. In: *2018 IEEE intelligent vehicles symposium (iv)*. IEEE, pp. 1051–1056.
- Fu, Yuan Yuan, Yang, Guijun, Ruiliang, Pu., et al., 2021. An overview of crop nitrogen status assessment using hyperspectral remote sensing: Current status and perspectives. *Eur. J. Agron.* 124, 126241.
- Gillespie, Alison (2022). PNOAA forecasts summer 'dead zone' of nearly 5.4K square miles in Gulf of Mexico. <https://www.noaa.gov/news-guy-release-noaa-forecasts-summer-dead-zone-of-nearly-54k-square-miles-in-gulf-of-mexico>. Accessed: 2022-09-07.
- Guyon, Isabelle, Weston, Jason, Barnhill, Stephen, et al., 2002. Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46 (1), 389–422.
- Hearst, Marti A., Susan T Dumais, Edgar Osuna, et al. (1998). "Support vector machines". In: *IEEE Intelligent Systems and their applications* 13.4, pp. 18–28.
- Hosmer, David W., Lemeshow, Stanley, Sturdivant, Rodney X., 2013. *Applied logistic regression*, Vol. 398. John Wiley & Sons.
- Ironmonger, Rebecca (2022). Rules on spreading slurry - does the EA's new approach provide a reprieve for farmers? <https://www.roythorne.co.uk/site/blog/agricultural-blog/rules-on-spreading-slurry>. Accessed: 2022-10-17.
- Jaihuni, Mustafa, Khan, Fawad, Lee, Deoghyun, et al., 2021. Determining Spatiotemporal Distribution of Macronutrients in a Cornfield Using Remote Sensing and a Deep Learning Model. *IEEE Access* 9, 30256–30266.
- Karanam, Shashmi (2021). Curse of Dimensionality — A "Curse" to Machine Learning. <https://towardsdatascience.com/curse-of-dimensionality-a-curse-to-machine-learning-c122ee33bfeb>. Accessed: 2022-10-17.
- Kirchman, David (2022). Dead zones: growing areas of aquatic hypoxia are threatening our oceans and rivers. <https://blog.oup.com/2021/Klec02/dead-zones-growing-areas-of-aquatic-hypoxia-are-threatening-our-oceans-and-rivers/>. Accessed: 2022-09-07.
- Klecka, William R., 1980. *Discriminant analysis*, Vol. 19. Sage.
- Kleinman, Peter J.A., Spiegel, Sheri, Liu, Jian, et al., 2020. Managing animal manure to minimize phosphorus losses from land to water. *Animal manure: Production, characteristics, environmental concerns, and management* 67, 201–228.
- Kursa, Miron B., Rudnicki, Witold R., 2010. Feature selection with the Boruta package. *J. Stat. Softw.* 36, 1–13.
- Liu, Jian, Kleinman, Peter JA, Aronsson, Helena, et al., 2018. A review of regulations and guidelines related to winter manure application. *Ambio* 47 (6), 657–670.
- Ma, Qiang, Wantai Yu, and Hua Zhou (2010). "The relationship between soil nutrient properties and remote sensing indices in the Phaeozem region of Northeast China". In: *2010 Second International Conference on Computational Intelligence and Natural Computing*, Vol. 2. IEEE, pp. 109–112.
- Mutanga, Onesimo, Dube, Timothy, Galal, Omer, 2017. Remote sensing of crop health for food security in Africa: Potentials and constraints. *Remote Sensing Appl.: Soc. Environ.* 8, 231–239.
- Orynbaiqzyy, Aiyun, Ursula Gessner, and Christopher Conrad (2019). "Crop type classification using a combination of optical and radar remote sensing data: A review". In: *international journal of remote sensing* 40.17, pp. 6553–6595.
- Pedrayes, Oscar D, Lema, Dario G, Garcia, Daniel F, et al., 2021. Evaluation of Semantic Segmentation Methods for Land Use with Spectral Imaging Using Sentinel-2 and PNOA Imagery. *Remote Sensing* 13 (12), 2292.
- Pereira, F.R. da S, J.P. de Lima, R.G. Freitas, et al. (2022). "Nitrogen variability assessment of pasture fields under an integrated crop-livestock system using UAV, PlanetScope, and Sentinel-2 data". In: *Computers and Electronics in Agriculture* 193, p. 106645.
- Quinlan, J. Ross, 1986. Induction of decision trees. *Mach. Learn.* 1, 81–106.
- Qun'ou, Jiang, Lidan, Xu., Siyang, Sun, et al., 2021. Retrieval model for total nitrogen concentration based on UAV hyper spectral remote sensing data and machine learning algorithms—A case study in the Miyun Reservoir, China. *Ecol. Ind.* 124, 107356.
- Romanko, Matthew (2017). "Remote Sensing in Precision Agriculture: Monitoring Plant Chlorophyll, and Soil Ammonia, Nitrate, and Phosphate in Corn and Soybean Fields". PhD thesis. Bowling Green State University.
- Sentinel-Hub (2022). Sentinel-2 RS indices — Sentinel-Hub custom scripts. <https://custom-scripts.sentinel-hub.com/custom-Serscripts/sentinel-2/indexdb/>. Accessed: 2023-04-03.
- Serra, Jean (1982). "Image analysis and mathematical morphology". In: (No Title).
- Shanmugapriya, P., Rathika, S., Ramesh, T., et al., 2019. Applications of remote sensing in agriculture—A Review. *Int. J. Current Microbiol. Appl. Sci.* 8 (1), 2270–2283.
- Shea, Kelly, Schaffer-Smith, Danica, Muenich, Rebecca L., 2022. Using remote sensing to identify liquid manure applications in eastern North Carolina. *J. Environ. Manage.* 317, 115334.
- Shou, Lina, Jia, Liangliang, Cui, Zhenling, et al., 2007. Using high-resolution satellite imaging to evaluate nitrogen status of winter wheat. *J. Plant Nutrition* 30 (10), 1669–1680.
- Sishodia, Rajendra P., Ray, Ram L., Singh, Sudhir K., 2020. Applications of remote sensing in precision agriculture: A review. *Remote Sensing* 12 (19), 3136.
- Tong, Yubing, Jayaram K Udupa, Chuang Wang, et al. (2018). "Radiomics-guided therapy for bladder cancer: Using an optimal biomarker approach to determine extent of bladder cancer invasion from t2-weighted magnetic resonance images". In: *Advances in Radiation Oncology* 3.3, pp. 331–338.
- Tzilizakis, John, Warner, D.J., Green, Andrew, et al., 2021. A broad-scale spatial analysis of the environmental benefits of fertiliser closed periods implemented under the Nitrates Directive in Europe. *J. Environ. Manage.* 299, 113674.
- Wang, Jingjing (2009). "Satellite Mapping of Past Biosolids (Sewage Sludge) and Animal Manure Application to Agriculture Fields in Wood County, Ohio". PhD thesis. Bowling Green State University.
- Ward, Joe H., 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58 (301), 236–244.
- Wold, Svante, Kim Esbensen, and Paul Geladi (1987). "Principal component analysis". In: *Chemometrics and intelligent laboratory systems* 2.1–3, pp. 37–52.
- Yang, Chun-Chieh, Prasher, Shiv O, Whalen, Joann, et al., 2002. PA—precision agriculture: use of hyperspectral imagery for identification of different fertilisation methods with decision-tree technology. *Biosyst. Eng.* 83 (3), 291–298.
- Zhu, Wanxue, Rezaei, Ehsan Eyshi, Nouri, Hamideh, et al., 2021. Quick detection of field-scale soil comprehensive attributes via the integration of UAV and sentinel-2B remote sensing data. *Remote Sensing* 13 (22), 4716.
- Zupan, Jure, 1994. Introduction to artificial neural network (ANN) methods: what they are and how to use them. *Acta Chim. Slov.* 41, 327–327.